311 Project Q2

William Clayton

November 2024

1. Log likelihood derivation.

   We have the probability of a question j being correctly answered by student i given by:

$$P(c_{ij} = 1|\theta_i, \beta_j) = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)}$$

   Thus, the probability for an incorrect answer is:

$$P(c_{ij} = 0|\theta_i, \beta_j) = 1 - \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)} = \frac{1}{1 + \exp(\theta_i - \beta_j)}$$

   Now we can derive the likelihood over all the data then take the log of this to get:

$$\log P(\boldsymbol{C}|\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{i,j} c_{ij} \log P(c_{ij} = 1|\theta_i, \beta_j) + (1 - c_{ij}) \log(P(c_{ij} = 0|\theta_i, \beta_j))$$

   Substituting the formulas, we get the log-likelihood for all students and questions:

$$\log P(\boldsymbol{C}|\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{i,j} [c_{ij}(\theta_i - \beta_j) - \log(1 + \exp(\theta_i - \beta_j))]$$

   Now, we can take the derivative of the log-likelihood with respect to θ_i to get:

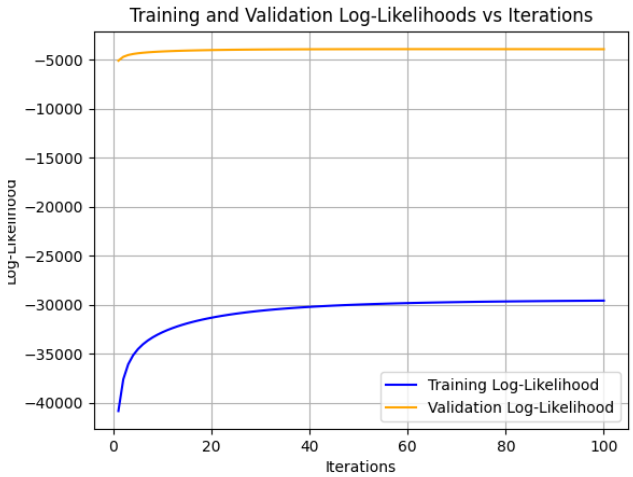
$$\frac{\partial}{\partial \theta_i} \log P(\boldsymbol{C}|\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_j c_{ij} - \sum_j \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)}$$

   Similarly, we can calculate the derivative with respect to β_j:

$$\frac{\partial}{\partial \beta_j} \log P(\boldsymbol{C}|\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_i \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)} - \sum_i c_{ij}$$

2. After experimenting with the model, the hyperparameters I chose are: learning rate $\alpha = 0.005$ with 100 iterations.

   Fig. 1: log-likelihoods for training and validation sets as a function of iteration.

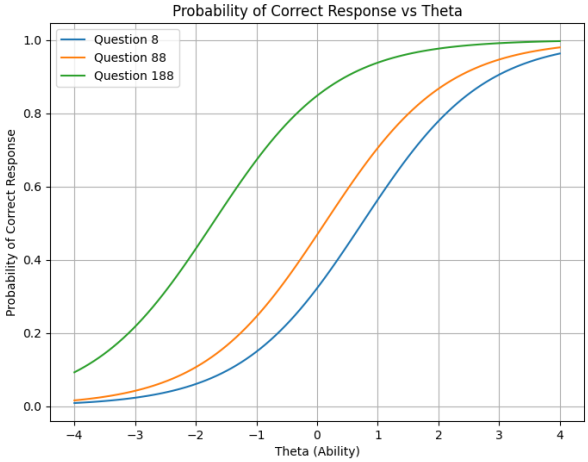


3. The final validation set accuracy is $0.706 = 70.6\%$, and the final test set accuracy is $0.708 = 70.8\%$

4. I chose $j_1 = 8, j_2 = 88, j_3 = 188$. The difficulty levels of each question are $\beta_8 = 0.744, \beta_{88} = 0.128, \beta_{188} = -1.717$.

   Fig. 2: Probability of getting the answer correct vs theta for each chosen question.

   These three curves represent the probability of a student with ability θ_i getting the correct answer for the corresponding question. We can see that the curves have a sigmoid shape, with the probability of a correct answer monotonically increasing with the ability (theta value) of the student. As the theta value approaches $-\infty$, the probability approaches 0.0 and as theta approaches ∞, the probability approaches 1.0. Notice that for an easier question like question 188 ($\beta_{188} = -1.717$), the probability curve lies above that of a harder question like question 8 ($\beta_8 = 0.744$). We can also see that when $\theta_i = \beta_j$, the probability of a correct answer is 0.5.


Probability of Correct Response vs Theta
Question 8
Question 88
Question 188
Probability of Correct Response
Theta (Ability)
1.0
0.8
0.6
0.4
0.2
0.0
−4
−3
−2
−1
0
1
2
3
4